Modelling the effects of air pollution on health using Bayesian Dynamic Generalised Linear Models

Duncan Lee (1) and Gavin Shaddick (2)
((1) University of Glasgow, (2) University of Bath)

February 2, 2008

Address for correspondence: Duncan Lee, Department of Statistics, 15 University Gardens, University of Glasgow, G12 8QQ, E-mail:duncan@stats.gla.ac.uk

Abstract

The relationship between short-term exposure to air pollution and mortality or morbidity has been the subject of much recent research, in which the standard method of analysis uses Poisson linear or additive models. In this paper we use a Bayesian dynamic generalised linear model (DGLM) to estimate this relationship, which allows the standard linear or additive model to be extended in two ways: (i) the long-term trend and temporal correlation present in the health data can be modelled by an autoregressive process rather than a smooth function of calendar time; (ii) the effects of air pollution are allowed to evolve over time. The efficacy of these two extensions are investigated by applying a series of dynamic and non-dynamic models to air pollution and mortality data from Greater London. A Bayesian approach is taken throughout, and a Markov chain monte carlo simulation algorithm is presented for inference. An alternative likelihood based analysis is also presented, in order to allow a direct comparison with the only previous analysis of air pollution and health data using a DGLM.

Key words *dynamic generalised linear model, Bayesian analysis, Markov chain monte carlo simulation, air pollution*

1 Introduction

The detrimental health effects associated with short-term exposure to air pollution is a major issue in public health, and the subject has received a great deal of attention in recent years. A number of epidemiological studies have found positive associations between common pollutants, such as particulate matter (measured as PM_{10}), ozone or carbon monoxide, and mortality or morbidity, with many of these associations relating to pollution levels below existing guidelines and standards (see, for example, Dominici et al. (2002), Vedal et al. (2003) or Roberts (2004)). These associations have been estimated from single-site and multi-city studies, the latter of which include 'Air pollution and health: a European approach' (APHEA) (Zmirou et al. (1998)) and 'The National Morbidity, Mortality, and Air Pollution Study' (NMMAPS) (Samet et al. (2000)). Although these studies have been conducted throughout the world in a variety of climates, positive associations have been consistently observed. The majority of these associations have been estimated using time series regression methods, and as the health data are only available as daily counts, Poisson generalised linear and additive models are the standard method of analysis. These data relate to the number of mortality or morbidity events that arise from the population living within a fixed region, for example a city, and are collected at daily intervals. Denoting the number of health events on day t by y_t, the standard log-linear model is given by

$$
\begin{aligned}
y_t &\sim \text{Poisson}(\mu_t) \quad \text{for } t = 1, \ldots, n, \\
\ln(\mu_t) &= w_t\gamma + \mathbf{z}_t^{\mathrm{T}}\boldsymbol{\alpha},
\end{aligned} \tag{1}
$$

in which the natural log of the expected health counts is linearly related to air pollution levels w_t and a vector of r covariates, $\mathbf{z}_t^{\mathrm{T}} = (z_{t1}, \ldots, z_{tr})$. The covariates model any seasonal variation, long-term trends, and temporal correlation present in the health data, and typically include smooth functions of calendar time and meteorological variables, such as temperature. If the smooth functions are estimated parametrically using regression splines, the model is linear, where as non-parametric estimation using smoothing splines, leads to an additive model.

In this paper we investigate the efficacy of using Bayesian dynamic generalised linear models (DGLMs, West et al. (1985) and Fahrmeir and Tutz (2001)) to analyse air pollution and health data. Dynamic generalised linear models extend generalised linear models by allowing the regression parameters to evolve over time via an autoregressive process of order p, denoted AR(p). The autoregressive nature of such models suggest two changes from the standard model (1) described above. Firstly, long-term trends and temporal correlation, present in the health data, can be modelled with an autoregressive process, which is in contrast to the standard approach of using a smooth function of calendar time. Secondly, the effects of air pollution can be modelled with an autoregressive process, which allows these effects to evolve over time. This evolution may be due to a change in the composition of individual pollutants, or because of a seasonal interaction with temperature. This is a comparatively new area of research, for which Peng et al. (2005) and Chiogna and Gaetan (2002) are the only known studies in this setting. The first of these forces the effects to follow a fixed seasonal pattern, which does not allow any other temporal

variation, such as a long-term trend. In contrast, Chiogna and Gaetan (2002) model this evolution as a first order random walk, which does not fix the temporal shape *a-priori*, allowing it to be estimated from the data. Their work is the only known analysis of air pollution and health data using DGLMs, and they implement their model in a likelihood framework using the Kalman filter. In this paper we present a Bayesian analysis based on Markov chain monte carlo (MCMC) simulation, which we believe is a more natural framework in which to analyse hierarchical models of this type.

The remainder of this paper is organised as follows. Section 2 introduces the Bayesian DGLM proposed here, and compares it to the likelihood based approach used by Chiogna and Gaetan (2002). Section 3 describes a Markov chain monte carlo estimation algorithm for the proposed model, while section 4 discusses the advantages of dynamic models for these data in more detail. Section 5 presents a case study, which investigates the utility of dynamic models in this context by analysing data from Greater London. Finally, section 6 gives a concluding discussion and suggests areas for future work.

2 Bayesian Dynamic generalised linear models

A Bayesian dynamic generalised linear model extends a generalised linear model by allowing a subset of the regression parameters to evolve over time as an autoregressive process. The general model proposed here begins with a Poisson assumption for the health data and is given by

$$\begin{aligned}
y_t &\sim \text{Poisson}(\mu_t) \qquad \text{for} \quad t=1,\ldots,n, \\
\ln(\mu_t) &= \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_t + \mathbf{z}_t^{\mathrm{T}}\boldsymbol{\alpha}, \\
\boldsymbol{\beta}_t &= F_1\boldsymbol{\beta}_{t-1} + \ldots + F_p\boldsymbol{\beta}_{t-p} + \boldsymbol{\nu}_t \qquad \boldsymbol{\nu}_t \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta), \\
\boldsymbol{\beta}_0,\ldots,\boldsymbol{\beta}_{-p+1} &\sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0), \\
\boldsymbol{\alpha} &\sim \mathrm{N}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\
\Sigma_\beta &\sim \text{Inverse-Wishart}(n_\Sigma, S_\Sigma^{-1}).
\end{aligned} \tag{2}$$

The vector of health counts are denoted by $\mathbf{y} = (y_1,\ldots,y_n)_{n\times 1}^{\mathrm{T}}$, and the covariates include an $r \times 1$ vector $\mathbf{z}_t$, with fixed parameters $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_r)_{r\times 1}^{\mathrm{T}}$, and a $q \times 1$ vector $\mathbf{x}_t$, with dynamic parameters $\boldsymbol{\beta}_t = (\beta_{t1},\ldots,\beta_{tq})_{q\times 1}^{\mathrm{T}}$. The dynamic parameters are assigned an autoregressive prior of order p, which is initialised by starting parameters $(\boldsymbol{\beta}_{-p+1},\ldots,\boldsymbol{\beta}_0)$ at times $(-p+1,\ldots,0)$. Each initialising parameter has a Gaussian prior with mean $\boldsymbol{\mu}_{0_{q\times 1}}$ and variance $\Sigma_{0_{q\times q}}$, and are included to allow $\boldsymbol{\beta}_1$ to follow an autoregressive process. The autoregressive parameters can be stacked into a single vector denoted by $\boldsymbol{\beta} = (\boldsymbol{\beta}_{-p+1},\ldots,\boldsymbol{\beta}_0,\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_n)_{(n+p)q\times 1}^{\mathrm{T}}$, and the variability in the process is controlled by a $q \times q$ variance matrix Σ_β, which is assigned a conjugate inverse-Wishart prior. For univariate processes Σ_β is scalar, and the conjugate prior simplifies to an inverse-gamma distribution. The evolution and stationarity of this process are determined by Σ_β and the $q \times q$ autoregressive matrices $F = \{F_1,\ldots,F_p\}$, the latter of which may contain unknown parameters or known constants, and the prior specification depends on its form. For example, a univariate

first order autoregressive process is stationary if $|F_1| < 1$, and a prior specification is discussed in section 3.1.4. A Gaussian prior is assigned to $\boldsymbol{\alpha}$ because prior information is simple to specify in this form. The unknown parameters are $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta)$ and components of, F, whereas the hyperparameters $(\boldsymbol{\mu}_\alpha, \Sigma_\alpha, n_\Sigma, S_\Sigma, \boldsymbol{\mu}_0, \Sigma_0)$ are known.

2.1 Estimation for a DGLM

We propose a Bayesian implementation of (2) using MCMC simulation, because it provides a natural framework for inference in hierarchical models. However, numerous alternative approaches have also been suggested, and a brief review is given here. West et al. (1985) proposed an approximate Bayesian analysis based on relaxing the normality of the AR(1) process, and assuming conjugacy between the data model and the AR(1) parameter model. They use Linear Bayes methods to estimate the conditional moments of $\boldsymbol{\beta}_t$, while estimation of Σ_β is circumvented using the discount method (Ameen and Harrison (1985)). Fahrmeir and co-workers (see Fahrmeir and Kaufmann (1991), Fahrmeir (1992) and Fahrmeir and Wagenpfeil (1997)) propose a likelihood based approach, which maximises the joint likelihood $f(\boldsymbol{\beta}|\mathbf{y})$. They use an iterative algorithm that simultaneously updates $\boldsymbol{\beta}$ and Σ_β, using the iteratively re-weighted Kalman filter and smoother, and expectation-maximisation (EM) algorithm (or generalised cross validation). This is the estimation approach taken by Chiogna and Gaetan (2002), and a comparison with our Bayesian implementation is given below. Other approaches to estimation include approximating the posterior density by piecewise linear functions (Kitagawa (1987)), using numerical integration methods (Fruhwirth-Schnatter (1994)), and particle filters Kitagawa (1996).

2.2 Comparison with the likelihood based approach

The main difference between this work and that of Chiogna and Gaetan (2002), who also used a DGLM in this setting, is the approach taken to estimation and inference. We propose a Bayesian approach with analysis based on MCMC simulation, which we believe has a number of advantages over the likelihood based analysis used by Chiogna and Gaetan. In a Bayesian approach the posterior distribution of $\boldsymbol{\beta}$ correctly allows for the variability in the hyperparameters, while confidence intervals calculated in a likelihood analysis do not. In a likelihood analysis (Σ_β, F) are estimated by data driven criteria, such as generalised cross validation, and estimates and standard errors of $\boldsymbol{\beta}$ are calculated assuming (Σ_β, F) are fixed at their estimated values. As a result, confidence intervals for $\boldsymbol{\beta}$ are likely to be too narrow, which may lead to a statistically insignificant effect of air pollution appearing to be significant. In contrast, the Bayesian credible intervals are the correct width, because $(\boldsymbol{\beta}, \Sigma_\beta, F)$ are simultaneously estimated within the MCMC algorithm.

The Bayesian approach allows the investigator to incorporate prior knowledge of the parameters into the model, whilst results similar to a likelihood analysis can be obtained by specifying prior ignorance. This is particularly important in dynamic models because the regression parameters are likely to evolve smoothly over time, and a non-informative prior for Σ_β may result in the estimated parameter process being contaminated with unwanted noise. Such noise may hide a trend in the parameter process, and can be removed by specifying an informative prior for Σ_β. The Bayesian approach is the natural framework in which to view hierarchical models

of this type, because it can incorporate variation at multiple levels in a straightforward manner, whilst making use of standard estimation techniques. In addition, the full posterior distribution can be calculated, whereas in a likelihood analysis only the mode and variance are estimated. However, as with any Bayesian analysis computation of the posterior distribution is time consuming, and likelihood based estimation is quicker to implement. To assess the relative performance of the two approaches, we apply all models in section 5 using the Bayesian algorithm described in section 3 and the likelihood based alternative used by Chiogna and Gaetan (2002).

The model proposed above is a re-formulation of that used by Chiogna and Gaetan (shown in equation (3) below), which fits naturally within the Bayesian framework adopted here. Apart from the inclusion of prior distributions in the Bayesian approach, there are two major differences between the two models, the first of which is operational and the second is notational. Firstly, a vector of covariates with fixed parameters ($\boldsymbol{\alpha}$) is explicitly included in the linear predictor, which allows the fixed and dynamic parameters to be updated separately in the MCMC simulation algorithm. This enables the autoregressive characteristics of $\boldsymbol{\beta}$ to be incorporated into its Metropolis-Hastings step, without forcing the same autoregressive property onto the simulation of the fixed parameters. This would not be possible in (3) as covariates with fixed parameters are included in the AR(1) process by a particular specification of Σ_β and F (diagonal elements of Σ_β are zero and F are one). This specification is also inefficient because n copies of each fixed parameter are estimated. Secondly, at first sight (3) appears to be an AR(1) process which compares with our more general AR(p) process. In fact an AR(p) process can be written in the form of (3) by a particular specification of $(\boldsymbol{\beta}, \Sigma_\beta, F)$, but we believe the approach given here is notationally clearer. In the next section we present an MCMC simulation algorithm for carrying out inference within this Bayesian dynamic generalised linear model.

$$\begin{aligned} y_t &\sim \text{Poisson}(\mu_t) & \text{for } \; t=1,\ldots,n \\ \ln(\mu_t) &= \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta}_t \\ \beta_t &= F\boldsymbol{\beta}_{t-1} + \boldsymbol{\nu}_t & \boldsymbol{\nu}_t \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta) \\ \boldsymbol{\beta}_0 &\sim \mathrm{N}(\boldsymbol{\mu}_0, \Sigma_0) \end{aligned} \tag{3}$$

3 MCMC estimation algorithm

The joint posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta, F)$ in (2) is given by

$$\begin{aligned} f(\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta, F|\mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta, F)f(\boldsymbol{\alpha})f(\boldsymbol{\beta}|\Sigma_\beta, F)f(\Sigma_\beta)f(F) \\ &= \prod_{t=1}^{n}\text{Poisson}(y_t|\boldsymbol{\beta}_t, \boldsymbol{\alpha})\mathrm{N}(\boldsymbol{\alpha}|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)\prod_{t=1}^{n}\mathrm{N}(\boldsymbol{\beta}_t|F_1\boldsymbol{\beta}_{t-1} + \ldots + F_p\boldsymbol{\beta}_{t-p}, \Sigma_\beta) \\ &\times \mathrm{N}(\boldsymbol{\beta}_{-p+1}|\boldsymbol{\mu}_0, \Sigma_0)\ldots\mathrm{N}(\boldsymbol{\beta}_0|\boldsymbol{\mu}_0, \Sigma_0)\text{Inverse-Wishart}(\Sigma_\beta|n_Q, S_Q^{-1})f(F), \end{aligned}$$

where $f(F)$ depends on the form of the AR(p) process. The next section describes the overall simulation algorithm, with specific details given in 3.1.1 - 3.1.4.

3.1 Overall simulation algorithm

The parameters are updated using a block Metropolis-Hastings algorithm, in which starting values $(\boldsymbol{\beta}^{(0)}, \boldsymbol{\alpha}^{(0)}, \Sigma_\beta^{(0)}, F^{(0)})$ are generated from overdispersed versions of the priors (for example t-distributions replacing Gaussian distributions). The parameters are alternately sampled from their full conditional distributions in the following blocks.

(a) **Dynamic parameters** $\boldsymbol{\beta} = (\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_n)$.
Further details are given in Section (3.1.1).

(b) **Fixed parameters** $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$.
Further details are given in Section (3.1.2).

(c) **Variance matrix** Σ_β.
Further details are given in section (3.1.3).

(d) **AR**(p) **matrices**, $F = (F_1, \ldots, F_p)$ (or components of).
Further details are given in section (3.1.4).

3.1.1 Sampling from $f(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \Sigma_\beta, F)$

The full conditional of $\boldsymbol{\beta}$ is the product of n Poisson observations and a Gaussian AR(p) prior given by

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\alpha}, \Sigma_\beta, F) &\propto \prod_{t=1}^{n} \text{Poisson}(y_t|\boldsymbol{\beta}_t, \boldsymbol{\alpha}) \prod_{t=1}^{n} \text{N}(\boldsymbol{\beta}_t|F_1\boldsymbol{\beta}_{t-1} + \ldots + F_p\boldsymbol{\beta}_{t-p}, \Sigma_\beta) \\ &\times \text{N}(\boldsymbol{\beta}_{-p+1}|\boldsymbol{\mu}_0, \Sigma_0) \cdots \text{N}(\boldsymbol{\beta}_0|\boldsymbol{\mu}_0, \Sigma_0). \end{aligned}$$

The full conditional is non-standard, and a number of simulation algorithms have been proposed that take into account the autoregressive nature of $\boldsymbol{\beta}$. Fahrmeir et al. (1992) combine a rejection sampling algorithm with a Gibbs step, but report acceptance rates that are very low making the algorithm prohibitively slow. In contrast, Shephard and Pitt (1997) and Gamerman (1998) suggest Metropolis-Hastings algorithms, in which the proposal distributions are based on Fisher scoring steps and Taylor expansions respectively. However, such proposal distributions are computationally expensive to calculate, and the conditional prior proposal algorithm of Knorr-Held (1999) is used instead. His proposal distribution is computationally cheap to calculate, compared with those of Shephard and Pitt (1997) and Gamerman (1998), while the Metropolis-Hastings acceptance rate has a simple form and is easy to calculate. Further details are given in appendix A.

3.1.2 Sampling from $f(\boldsymbol{\alpha}|\mathbf{y}, \boldsymbol{\beta}, \Sigma_\beta, F)$

The full conditional of $\boldsymbol{\alpha}$ is non-standard because it is the product of a Gaussian prior and n Poisson observations. As a result, simulation is carried out using a Metropolis-Hastings step, and two common choices are random walk and Fisher scoring proposals (for details see Fahrmeir and Tutz (2001)). A random walk proposal is used here because of its computational cheapness compared with the Fisher scoring alternative, and the availability of a tuning parameter. The parameters are

updated in blocks, which is a compromise between the high acceptance rates obtained by univariate sampling, and the improved mixing that arises when large sets of parameters are sampled simultaneously. Proposals are drawn from a Gaussian distribution with mean equal to the current value of the block and a diagonal variance matrix. The diagonal variances are typically identical and can be tuned to give good acceptance rates.

3.1.3 Sampling from $f(\Sigma_\beta|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, F)$

The full conditional of Σ_β comprises n AR(p) Gaussian distributions for $\boldsymbol{\beta}_t$ and a conjugate inverse-Wishart(n_Σ, S_Σ^{-1}) prior, which results in an inverse-Wishart(a, b) posterior distribution with

$$\begin{aligned} a &= n_\Sigma + n, \\ b &= \left(S_\Sigma + \sum_{t=1}^{n}(\boldsymbol{\beta}_t - F_1\boldsymbol{\beta}_{t-1} - \ldots - F_p\boldsymbol{\beta}_{t-p})(\boldsymbol{\beta}_t - F_1\boldsymbol{\beta}_{t-1} - \ldots - F_p\boldsymbol{\beta}_{t-p})^T\right)^{-1}. \end{aligned}$$

However, the models applied in section five are based on univariate autoregressive processes, for which the conjugate prior simplifies to an inverse-gamma distribution. If a non-informative prior is required, an inverse-gamma(ϵ, ϵ) prior with small ϵ is typically used. However as discussed in section 2.2, an informative prior may be required for Σ_β, and representing informative prior beliefs using a member of the inverse-gamma family is not straightforward. The variance parameters of the autoregressive processes are likely to be close to zero (to ensure the evolution is smooth), so we represent our prior beliefs as a Gaussian distribution with zero mean, which is truncated to be positive. The informativeness of this prior is controlled by its variance, with smaller values resulting in a more informative distribution. If this prior is used, the full conditional can be sampled from using a Metropolis-Hastings step with a random walk proposal.

3.1.4 Sampling from $f(F|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \Sigma_\beta)$

The full conditional of F depends on the form and dimension of the AR(p) process, and the most common types are univariate AR(1) ($\beta_t \sim \text{N}(F_1\beta_{t-1}, \Sigma_\beta)$) and AR(2) ($\beta_t \sim \text{N}(F_1\beta_{t-1} + F_2\beta_{t-2}, \Sigma_\beta)$) processes. In either case, assigning (F_1) or (F_1, F_2) flat priors results in a Gaussian full conditional distribution. For example in a univariate AR(1) process, the full conditional for F_1 is Gaussian with mean $\frac{\sum_{t=1}^{n}\beta_t\beta_{t-1}}{\sum_{t=1}^{n}\beta_{t-1}^2}$, and variance $\frac{\Sigma_\beta}{\sum_{t=1}^{n}\beta_{t-1}^2}$. Similar results can be found for an AR(2) process.

4 Modelling air pollution and health data

As described in the introduction, air pollution and health data are typically modelled by Poisson linear or additive models, which are similar to equation (1). The daily health counts are regressed against air pollution levels and a vector of covariates, the latter of which model long-term trends, seasonal variation and temporal correlation commonly present in the daily mortality series. The covariates typically include an intercept term, indicator variables for day of the week, and smooth functions of

calendar time and meteorological covariates, such as temperature. A large part of the seasonal variation is modelled by the smooth function of temperature, while the long-term trends and temporal correlation are removed by the smooth function of calendar time. The air pollution component typically has the form $w_t\gamma$, which forces its effect on health to be constant. Analysing these data with dynamic models allows this standard approach to be extended in two ways, both of which are described below.

4.1 Modelling long-term trends and temporal correlation

The autoregressive nature of a dynamic generalised linear model, enables long-term trends and temporal correlation to be modelled by an autoregressive process, rather than a smooth function of calendar time. This is desirable because such a process sits in discrete time and estimates the underlying trend in the data $\{t, y_t\}_{t=1}^n$, while its smoothness is controlled by a single parameter (the evolution variance). In these respects an autoregressive process is a natural choice to model the influence of confounding factors because it can be seen as the discrete time analogue of a smooth function of calendar time. In the dynamic modelling literature (see for example Chatfield (1996) and Fahrmeir and Tutz (2001)), long-term trends are commonly modelled by:

$$\begin{aligned}
\text{First order random walk} \quad \beta_t &\sim \mathrm{N}(\beta_{t-1}, \tau^2),\\
\text{Second order random walk} \quad \beta_t &\sim \mathrm{N}(2\beta_{t-1} - \beta_{t-2}, \tau^2), \qquad (4)\\
\text{Local linear trend model} \quad \beta_t &\sim \mathrm{N}(\beta_{t-1} + \delta_{t-1}, \tau^2),\\
\delta_t &\sim \mathrm{N}(\delta_{t-1}, \psi^2).
\end{aligned}$$

All three processes are non-stationary which allows the underlying mean level to change over time, a desirable characteristic when modelling long-term trends. A second order random walk is the natural choice from the three alternatives, because it is the discrete time analogue of a natural cubic spline of calendar time (Fahrmeir and Tutz (2001)), one of the standard methods for estimating the smooth functions. Chiogna and Gaetan (2002) also use a second order random walk for this reason, but in section five we extend their work by comparing the relative performance of smooth functions and each of the three processes listed above. We estimate the smooth function with a natural cubic spline, because it is parametric, making estimation within a Bayesian setting straightforward.

4.2 Modelling the effects of air pollution

The effects of air pollution are typically assumed to be constant (represented by γ), or depend on the level of air pollution, the latter of which replaces $w_t\gamma$ in (1) with a smooth function $f(w_t|\lambda)$. This is called a dose-response relationship, and higher pollution levels typically result in larger adverse effects. Comparatively little research has allowed these effects to evolve over time, and any temporal variation is likely to

be seasonal or exhibit a long-term trend. Seasonal effects may be caused by an interaction with temperature, or with another pollutant exhibiting a seasonal pattern. In contrast, long-term trends may result from a slow change in the composition of harmful pollutants, or from a change in the size and structure of the population at risk over a number of years. The only previous studies which investigate the time-varying effects of air pollution are those of Peng et al. (2005) and Chiogna and Gaetan (2002), who model the temporal evolution as $\gamma_t = \theta_0 + \theta_1 \sin(2\pi t/365) + \theta_2 \cos(2\pi t/365)$ and $\gamma_t \sim \mathrm{N}(\gamma_{t-1}, \sigma^2)$ respectively. The seasonal form is restrictive because it does not allow the temporal variation to exhibit shapes which are not seasonal. In contrast, the first order random walk used by Chiogna and Gaetan (2002) does not fix the form of the time-varying effects *a-priori*, allowing their shape to be estimated from the data, which results in a more realistic model. In section five we also model this temporal variation with a first order random walk, because of its flexibility and because it allows a comparison with the work of Chiogna and Gaetan (2002).

5 Case study analysing data from Greater London

The extensions to the standard model described in section 4 are investigated by analysing data from Greater London. The first subsection describes the data that are used in this case study, the second discusses the choice of statistical models, while the third presents the results.

5.1 Description of the data

The data used in this case study relate to daily observations from the Greater London area during the period 1^{st} January 1995 until 31^{st} December 1997. The health data comprise daily counts of respiratory mortality drawn from the population living within Greater London, and are shown in Figure 1. A strong seasonal pattern is evident, with a large increase in the number of deaths during the winter of 1996/1997. The cause of this peak is unknown, and research has shown no influenza epidemic during this time (which has previously been associated with large increases of this type (Griffin and Neuzil (2002)). The air pollution data comprise particulate matter levels, which are measured as PM_{10} at eleven locations across Greater London. To obtain a single measure of PM_{10}, the values are averaged across the locations, a strategy which is commonly used in studies of this type(see for example Katsouyanni et al. (1996) and Samet et al. (2000)). For these data this strategy is likely to introduce minimal additional exposure error, because PM_{10} levels in London between 1994 and 1997 exhibit little spatial variation (Shaddick and Wakefield (2002)). In addition to the health and pollution data, a number of meteorological covariates including indices of temperature, rainfall, wind speed and sunshine, are measured at Heathrow airport. However, in this study only daily mean temperature, measured in Celsius (^{0}C), is a significant covariate and the rest are not used.

5.2 Description of the statistical models used

Dynamic generalised linear models extend the standard approach to analysing these data by: (i) allowing the trend and temporal correlation in the health data to be removed with an autoregressive process; (ii) allowing the effects of air pollution to

evolve over time. To investigate these two extensions eight models are applied to the Greater London data, and a summary is given in Table 1. The general form of all eight models is given by

$$\begin{aligned} y_t &\sim \text{Poisson}(\mu_t) \quad \text{for } t = 1, \ldots, n, \\ \ln(\mu_t) &= \text{PM}_{10_{t-1}}\gamma_t + \beta_t + S(\text{temperature}_t|3, \boldsymbol{\alpha}_3), \\ \boldsymbol{\alpha} &\sim \text{N}(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \end{aligned} \tag{5}$$

where β_t is the trend component, and γ_t represents the effect of PM_{10} on day t. The trend component is represented by one of four sub-models, denoted by (a) - (d) below, which take the form of a natural cubic spline of calendar time or one of the three autoregressive processes given in (4).

(a) Natural cubic splines

$\beta_t = \alpha_1 + S(t|k, \boldsymbol{\alpha}_2)$

(b) First order random walk

$\beta_t \sim \text{N}(\beta_{t-1}, \tau^2)$
$\beta_0 \sim \text{N}(3.5, 10)$
$\tau^2 \sim \text{N}(0, g_2)_{I[\tau^2>0]}$

(c) Second order random walk

$\beta_t \sim \text{N}(2\beta_{t-1} - \beta_{t-2}, \tau^2)$
$\beta_{-1}, \beta_0 \sim \text{N}(3.5, 10)$
$\tau^2 \sim \text{N}(0, g_2)_{I[\tau^2>0]}$

(d) Local linear trend

$\beta_t \sim \text{N}(\beta_{t-1} + \delta_{t-1}, \tau^2)$
$\delta_t \sim \text{N}(\delta_{t-1}, \psi^2)$
$\beta_0 \sim \text{N}(3.5, 10)$
$\delta_0 \sim \text{N}(0, 10)$
$\tau^2 \sim \text{N}(0, g_4)_{I[\tau^2>0]}$
$\psi^2 \sim \text{N}(0, g_5)_{I[\psi^2>0]}$

The effects of air pollution are represented by one of two components, denoted by (i) and (ii) below, which force these effects to be constant or allow them to evolve over time.

(i) Constant

$\gamma_t = \gamma$

(ii) Time-varying - first order random walk

$\gamma_t \sim \text{N}(\gamma_{t-1}, \sigma^2)$
$\gamma_0 \sim \text{N}(0, 10)$
$\sigma^2 \sim \text{N}(0, g_1)_{I[\sigma^2>0]}$

In the model description above, $\text{N}(0, g_1)_{I[\sigma^2>0]}$ denotes a truncated Gaussian distribution where $I[]$ is an indicator function which specifies the range of allowed (non-truncated) values. The smooth functions $S(var|df, \boldsymbol{\alpha})$ are estimated with natural cubic splines, where *var* is the covariate and *df* is the degrees of freedom. The vector of fixed parameters is different for each model, and includes the intercept, the parameters that make up the natural cubic splines, and the constant effect of air pollution. To compare the results with those presented by Chiogna and Gaetan (2002), each model is analysed within the Bayesian approach described here and their likelihood based alternative. Likelihood based analysis is carried out using

the iteratively re-weighted Kalman filter and smoother proposed by Fahrmeir and Wagenpfeil (1997), while the hyperparameters are estimated using Akaike Information Criterion (AIC). The remainder of this subsection describes the model building process, including justifications for the choice of models. The first part focuses on the trend models, while the second discusses the air pollution component.

5.2.1 Modelling trends, seasonal variation and temporal correlation

The model building process began by removing the trend, seasonal variation and temporal correlation from the respiratory mortality series. These data exhibit a pronounced yearly cycle, which is partly modelled by the trend component β_t, and partly by daily mean temperature (also has a yearly cycle). The latter was added to the model at a number of different lags with different shaped relationships, and the fit to the data was assessed using the deviance information criterion (DIC, Spiegelalter et al. (2002)). As a result, a smooth function of the same days temperature with three degrees of freedom is used in the final models, because it has the lowest DIC, and has previously been shown to have a U-shaped relationship with mortality (see for example Dominici et al. (2000)). The smooth function is modelled with a natural cubic spline, because it is fully parametric making analysis within a Bayesian setting straightforward.

The smooth function of calendar time (trend component (a)) is modelled by a natural cubic spline for the same reason, and has previously been used by Daniels et al. (2004)). The smoothness of the spline is chosen by DIC to be 27, and is fixed prior to analysis. To allow a fairer comparison with the other trend components, the degrees of freedom should be estimated simultaneously within the MCMC algorithm, but this makes the average trend impossible to estimate. As the smoothness of the spline is fixed, its parameters (part of $\boldsymbol{\alpha}$) are given a non-informative Gaussian prior. In the Likelihood analysis, the smoothing parameter is chosen by minimising AIC which also leads to 27 degrees of freedom.

The remaining three trend models are based on autoregressive processes, and their smoothness is controlled by the evolution variances (τ^2, ψ^2). Initially, these variances were assigned non-informative inverse-gamma$(0.01, 0.01)$ priors, but the estimated trends (not shown) just interpolates the data. This undesirable aspect can be removed by assigning (τ^2, ψ^2) informative priors, which shrink their estimates towards zero producing a smoother trend. The choice of an informative prior within the inverse-gamma family is not straightforward, and instead we represent our prior beliefs as a Gaussian distribution with mean zero, which is truncated to be positive. This choice of prior forces (τ^2, ψ^2) to be close to zero, with the prior variances, denoted by (g_2, g_3, g_4, g_5), controlling the level of informativeness. Smaller prior variances results in more prior weight close to zero, forcing the estimated process to be smoother.

It seems likely that the trend in mortality will be similar on consecutive days, meaning that the autoregressive process should evolve smoothly over time. The trend is modelled on the linear predictor scale, which corresponds to the natural log of the data and has a range between 2.5 and 4.5 daily deaths (between about 12 and

90 on the un-logged scale). On that scale, a jump of 0.01 on consecutive days is approximately the largest difference that cannot be detected by the eye, resulting in a visually smooth trend. To relate this to the choice of (g_2, g_3, g_4, g_5), each of the three processes were simulated with a variety of variances, and the average absolute difference between consecutive values was calculated. The variances were chosen so that 50% of the prior mass was below the threshold value that gave average differences of 0.01, resulting in $g_2 = 10^{-7}$, $g_3 = 10^{-14}$. The local linear trend model has two variance parameters, and it was found that both needed to be tightly controlled for the process to evolve smoothly, resulting in $g_4 = g_5 = 10^{-16}$. Sensitivity analyses were carried out for different values of (g_2, g_3, g_4, g_5), but it was found that larger values resulted in trends that were not visually smooth. In the likelihood analysis, the variance parameters are chosen by optimising AIC. The priors for the initialising parameters $(\beta_{-1}, \beta_0, \delta_0)$ are non-informative Gaussian distributions with mean equal to zero for the rate δ_0, and 3.5 for (β_{-1}, β_0), the average of mortality data from previous years on the logged scale.

5.2.2 Modelling the effects of PM_{10}

After modelling the influence of unmeasured risk factors, the effects of PM_{10} at a number of different lags were investigated. A lag of one day is used in the final models, because it has the minimum DIC and has been used in other recent studies (see for example Dominici et al. (2000), and Zhu et al. (2003)). Constant and time-varying effects of PM_{10} are investigated in this case study, with the latter modelled by a first order random walk, which allows a comparison with the work of Chiogna and Gaetan (2002). Initially, a non-informative inverse-gamma$(0.01, 0.01)$ prior was specified for the variance of the random walk (denoted by σ^2), but the estimated time-varying effects (not shown) are contaminated by noise and an underlying trend cannot be seen. These effects are likely to evolve smoothly over time, and to enforce this smoothness σ^2 is assigned an informative zero mean Gaussian prior which is truncated to be positive. The informativeness is controlled by the variance g_1, which is chosen using an identical approach to that described above. In this case the likely range of effects is -0.003 to 0.005, and the largest difference that is undetectable by the eye is around 0.00005, leading to $g_1 = 10^{-16}$. To corroborate this choice a sensitivity analysis was carried out for different values of g_1, which showed the evolution was smooth for values as large as 10^{-10}. As this is less informative than 10^{-16}, it is used in the final models. In the likelihood analysis, the variance parameter is chosen by optimising the AIC.

5.3 Results

The models contain a large number of parameters, so to aid convergence the covariates (PM_{10} and the basis functions for the natural cubic splines of calendar time and temperature) are standardised to have a mean of zero and a standard deviation of one before inclusion in the model (and are subsequently back-transformed when obtaining results from the posterior distribution). The Markov chains are burnt in for 40,000 iterations, by which point convergence was assessed to have been reached using the methods of Gelman et al. (2003). At this point a further 100,000 iterations are simulated, which are thinned by 5 to reduce autocorrelation, resulting in 20,000 samples from the joint posterior distribution.

5.3.1 Results for the four trend models β_t

Long-term trends, overdispersion and temporal correlation are removed from the health data with one of four trend models: a natural cubic spline of calendar time (models 1 and 2); a first order random walk (models 3 and 4); a second order random walk (models 5 and 6); and a local linear trend model (models 7 and 8). To aid clarity in the following discussion, these approaches are compared and contrasted assuming a constant effect of air pollution (using the odd numbered models). Figure 1 shows the health data from Greater London, together with the estimated trends from the Bayesian (solid lines) and likelihood (dotted lines) analyses. Panel (a) shows the estimated trend from a natural cubic spline of calendar time, panel (b) relates to the first order random walk, panel (c) to the second order random walk, and panel (d) to the local linear trend model. All four models capture the underlying trend in the health data well, and the Bayesian and likelihood estimates are very similar. The only major differences between the eight estimates are in the winters of 1996 and 1997, where the respiratory mortality data has yearly peaks. For each trend model, the Bayesian estimate captures the height of these peaks better than its likelihood counterpart, while the second order random walk outperforms the other three alternatives. For example, in the winter of 1997 the maximum number of deaths on a single day is 145, and the Bayesian estimates of this peak are, (a) - 98.4, (b) - 92.2, (c) - 107.6, (d) - 101.5, while the corresponding likelihood values are, (a) - 69.4, (b) - 79.1, (c) - 85.9, (d) - 85.1. These figures show that the second order random walk is the most adept at modelling these peaks, while the local linear trend model outperforms both the natural cubic spline and first order random walk.

All eight estimates have the same visual smoothness, and a summary of the smoothing parameters is given in Table 2. For the natural cubic spline model, which does not estimate the number of basis functions as part of the MCMC algorithm (it is estimated by DIC), the estimate of k is identical in both analyses (it is estimated by GCV in a likelihood analysis). However, for the remaining three analyses, the likelihood estimates of the smoothing parameters (τ^2) are significantly larger than their Bayesian counterparts, without the corresponding trends being less smooth. This is unexpected, and is most likely caused by differences in the techniques used to estimate the autoregressive processes, a point which is taken up in the discussion.

To examine how effective each trend model is at removing temporal correlation from the health data, a measure of the residuals is required. In a Bayesian setting residuals are not well defined (see Pettit (1986)), because there is no natural point estimate for the parameters. Instead, a 'residual distribution' can be generated for each y_t by simulation. For example, a Pearson type distribution has a jth 'realised residual'

$$r_t^{(j)} = \frac{y_t - \mathbb{E}[y_t|\boldsymbol{\theta}^{(j)}]}{\sqrt{\text{Var}[y_t|\boldsymbol{\theta}^{(j)}]}},$$

in which $\boldsymbol{\theta}^{(j)}$ is the jth sample from the joint posterior distribution. The residual distribution takes into account the uncertainty in $\boldsymbol{\theta}$, and residuals based on point estimates are approximations to this distribution. Figure 2 shows the autocorrelation

function of an approximation to this residual distribution, which is based on posterior medians. The second order random walk again outperforms the other approaches, showing little or no correlation in the approximate Pearson residuals. In contrast, the natural cubic spline is the worst of the four trend components, having significant correlation at the first four lags. The remaining two models perform similarly, and only show significant correlation at the first lag. The residuals from the likelihood analyses (not shown) show a similar comparison between the four approaches, but exhibit greater correlation than those from the Bayesian analysis, suggesting that the Bayesian models are superior. A plot of the residuals against time showed little difference between the four approaches and is not shown.

5.3.2 Results for the time-varying effects of air pollution γ_t

The models presented here allow the effects of air pollution to evolve over time as a first order random walk, or be fixed at a constant value. In the graphs and tables that follow, these effects are given as a relative risk for an increase in 10 units of PM_{10}. This is calculated as the ratio of expected number of deaths, μ_t^{+10}/μ_t, where μ_t^{+10} is the expected number of deaths if the air pollution level had risen by 10 units. The relative risk is given by $\exp(10\gamma)$ (for models 1,3,5 and 7) and a value of 1 represents no effect of air pollution.

(a) - Constant effects
Table 3 shows the estimated relative risks from models 1, 3, 5 and 7, which force the effects of air pollution to be constant. All eight Bayesian and likelihood estimates are very similar (range from 1.007 to 1.015), suggesting that the method of analysis and the choice of trend component do not affect the estimated health risk. The estimates from the likelihood analyses are always larger than those from the corresponding Bayesian model, although the differences are not large. The Bayesian credible intervals are wider than their likelihood counterparts, and few of the intervals contain one, suggesting that exposure to PM_{10} has a statistically significant effect on mortality.

(b) - Time-varying effects
In the likelihood analyses the estimated variance parameters are all zero, forcing the time-varying effect to be constant. However, in the Bayesian analyses these estimates are greater than zero, and the time-varying effects are shown in Figure 3. The evolution in the effects is smooth, which is a result of the informative prior placed on the variance σ^2. All four estimates are very similar, suggesting that the choice of model for the unmeasured risk factors does not affect the substantive conclusions. The effects exhibit a slowly increasing long-term trend, which has ranges of: (a) 1.005 to 1.015; (b) 1.007 to 1.014; (c) 1.002 to 1.019; (d) 0.999 to 1.024. The 95% credible intervals for panels (a) and (b) (models 2 and 4 respectively) are of a similar size, but the remaining two exhibit substantial additional variation, especially in panel (d). This additional variation is not supported by the same pattern in the credible intervals for the constant effects, a point which is taken up in the discussion. However, the width of the four intervals in panels (a) to (d) suggest that a constant

effect of PM_{10} cannot be ruled out.

6 Discussion

This paper proposes the use of Bayesian dynamic generalised linear models to estimate the relationship between air pollution exposure and mortality or morbidity. The majority of air pollution and health studies fix the effects of air pollution to be constant over time, and model long-term trends and temporal correlation in the health data using a smooth function of calendar time. The DGLM framework allows autoregressive processes to be used for both these factors, the first of which allows the effects of air pollution change over time. A Bayesian approach is assumed throughout with analysis based on MCMC simulation. In addition a likelihood analysis is also presented, which allows a comparison with the only previous air pollution and health study that used a DGLM.

The results from the four trend models lead to two main conclusions. Firstly, although all four trend components capture the underlying level of daily deaths relatively well, the standard approach of using smooth functions is outperformed by the autoregressive processes. In particular, the best of these processes is the second order random walk, because its residuals exhibit no correlation, and the two winter peaks in daily mortality are well represented. In contrast, the smooth function leaves significant correlation in the residuals, while the estimated peaks are captured less well. The local linear trend model also performs better than the smooth function, but the first order random walk gives similar results. The poor performance of the smooth function is most likely caused by the way it is estimated, which includes the choice of smoothing parameter and the use of natural cubic splines. The smooth function's degrees of freedom is estimated by DIC and is fixed during the simulation, which is in contrast to the autoregressive processes whose smoothing parameters are estimated within the MCMC algorithm. As a result, the estimated autoregressive trends incorporate the variation in their smoothing parameter, which is not the case for the smooth function and may account for the latter's poorer performance. Another possible cause of the smooth function's poorer performance is the use of natural cubic splines to estimate it. Regression splines were used here because of their parametric make-up, but are known to be less flexible than non-parametric alternatives. An interesting area of future research would be to compare the performances of the trend models used here, against non-parametric smooth functions, such as smoothing splines or LOESS smoothers.

Secondly, the Bayesian approach gives results that are superior to the likelihood analysis, both in terms of removing temporal correlation from the health data, and its ability to capture winter peaks in mortality. The estimated smoothing parameters for the Bayesian and likelihood implementations of the natural cubic splines are obtained by optimising data driven criteria (DIC and AIC), and it is not surprising that both estimates are identical. However, for the autoregressive processes the Bayesian estimates are smaller than their likelihood counterparts, which is caused by the relative strengths of the truncated Gaussian prior and the penalty term in the AIC criteria. A sensitivity analysis shows that such a strong prior is required, because using a non-informative prior for τ^2 results in the estimated trend interpolat-

ing the data. An initial comparison of the estimated smoothing parameters (τ^2) for the Bayesian and likelihood analyses, shows that the latter were larger and therefore might be expected to produce a trend exhibiting greater variability (and thus model the data at the peaks more accurately). However, the opposite was observed, and the larger estimates of τ^2 in the likelihood analyses result in trends which are less variable. This apparent anomaly is most likely caused by differences in the methods used to implement the autoregressive constraint for $\boldsymbol{\beta}$. In the Bayesian analysis, this is implemented through the specification of an autoregressive prior $f(\boldsymbol{\beta})$, whereas the likelihood approach enforces the autoregressive constraint using the Kalman filter. The filter uses a two stage process which firstly estimates $\mathbb{E}[\boldsymbol{\beta}_t|y_1,\ldots,y_t]$, for all t, and then smoothes the results by estimating $\mathbb{E}[\boldsymbol{\beta}_t|y_1,\ldots,y_n]$. The final likelihood estimates are based on these smoothed values, and it is this additional smoothing imposed by the Kalman filter, that reduces the variability in the estimated trends, which over smoothes the data in this case.

The Bayesian estimates of the pollution-mortality relationship exhibit a consistent long-term pattern regardless of the choice of trend model, suggesting that this temporal variation should be investigated further. However, no seasonal interaction is observed, meaning that the model of Peng et al. (2005) is too restrictive for these data. The informative prior for σ^2 forces these effects to evolve smoothly over time, while a sensitivity analysis showed that using a non-informative prior leads to the estimate being contaminated with noise. This noise is caused by the excess number of parameters used to model the time-varying effects, which makes these parameters non-identifiable. A non-informative prior for σ^2 is too weak for these data, and the specification of an informative prior shrinks the evolution variance towards zero, effectively reducing the number of parameters. The resulting temporal evolution is smooth, but this is achieved at the expense of a very informative prior.

The estimated time-varying effects are not altered by the choice of trend model, although the credible intervals increase in width if a second order random walk or local linear trend are used. These two represent the most flexible trend models, and their increased variation may cause slight non-identifiability or collinearity with the time-varying effects of PM_{10}, reducing their precision. The estimated temporal variation from the Bayesian models exhibit a similar shape to those reported by Chiogna and Gaetan (2002) in Birmingham Alabama, using a likelihood approach. However, this contrasts with our likelihood based analyses which forced the effects to be constant and not exhibit any temporal variation. The difference in curvature between our Bayesian and likelihood analyses is again due to the way the smoothing parameters are estimated. The likelihood approach calculates the likelihood for a range of values of the smoothing parameter, and estimates σ^2 by optimising a data driven criterion. In contrast, the Bayesian approach averages over the posterior for σ^2, which incorporates the possibility of no smoothing, thus leading to an estimate which exhibits greater curvature.

Acknowledgements

We would like to thank the Small Area Health Statistics Unit, which is funded by grants from the Department of Health, Department of the Environment, Food and

Rural Affairs, Health and Safety Executive, Scottish Executive, National Assembly of Wales, and Northern Ireland Assembly, who provided the mortality data from Greater London.

Appendix A - simulation of $\boldsymbol{\beta}$

The first p parameters are updated separately from $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$, because their full conditional distribution does not depend on $\mathbf{y}$ and is a standard Gaussian distribution. In contrast, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n$ are sampled using a block Metropolis-Hastings scheme, in which the proposal distribution is based on the autoregressive prior. Ignoring $\boldsymbol{\beta}_{-p+1}, \ldots, \boldsymbol{\beta}_0$ which have already been sampled, the autoregressive prior can be written as a singular multivariate Gaussian distribution

$$
\begin{aligned}
f(\boldsymbol{\beta}|F, \Sigma_\beta) &= \prod_{t=1}^{n} \mathrm{N}(\boldsymbol{\beta}_t | F_1\boldsymbol{\beta}_{t-1} + \ldots + F_p\boldsymbol{\beta}_{t-p}, \Sigma_\beta) \\
&\propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}^T K \boldsymbol{\beta}\right),
\end{aligned}
$$

with mean zero and singular precision matrix K. The precision matrix is given by

$$
K = \begin{bmatrix} K_{-p+1,-p+1} & \ldots & K_{-p+1,n} \\ \vdots & & \vdots \\ K_{n,-p+1} & \ldots & K_{n,n} \end{bmatrix}_{(n+p)q\times(n+p)q},
$$

where $K_{t,t}$ is a $q \times q$ block relating to $\boldsymbol{\beta}_t$. The blocks depend on the order of the AR(p) process, and K has a bandwidth of p blocks (all blocks K_{ij}, for which $|i-j| > p$ are zero). For example, an AR(1) process leads to

$$
\begin{aligned}
K_{t,t} &= \begin{cases} F_1^{\mathrm{T}}\Sigma_\beta^{-1}F_1 & t = 0 \\ F_1^{\mathrm{T}}\Sigma_\beta^{-1}F_1 + \Sigma_\beta^{-1} & t = 1, \ldots, n-1 \\ \Sigma_\beta^{-1} & t = n \end{cases}, \\
K_{t,t+1} &= -F_1^{\mathrm{T}}\Sigma_\beta^{-1} \qquad \forall\ t, \\
K_{t,t-1} &= -\Sigma_\beta^{-1}F_1 \qquad \forall\ t.
\end{aligned}
$$

The parameters are updated in blocks of size g, which is used as a tuning parameter to achieve the desired acceptance rates. The proposal distribution for a block $\boldsymbol{\beta}_{r,s} = (\boldsymbol{\beta}_r, \ldots, \boldsymbol{\beta}_s)_{gq\times 1}$, in which $s = r + g - 1$ is given by

$$
f(\boldsymbol{\beta}_{r,s}|\boldsymbol{\beta}_{-r,s}, F, \Sigma_\beta) \sim \mathrm{N}(\boldsymbol{\mu}_{r,s}, \Sigma_{r,s}),
$$

where $\boldsymbol{\beta}_{-r,s}$ denotes all elements of $\boldsymbol{\beta}$ except $\boldsymbol{\beta}_{r,s}$. The mean and variance are given by

$$
\begin{aligned}
\boldsymbol{\mu}_{r,s} &= \begin{bmatrix} -\tilde{K}_{r,s}^{-1}\tilde{K}_{-p+1,r-1}\boldsymbol{\beta}_{-p+1,r-1} & \text{if s=n} \\ -\tilde{K}_{r,s}^{-1}\tilde{K}_{s+1,n}\boldsymbol{\beta}_{s+1,n} & \text{if r=-p+1} \\ -\tilde{K}_{r,s}^{-1}(\tilde{K}_{-p+1,r-1}\boldsymbol{\beta}_{-p+1,r-1} + \tilde{K}_{s+1,n}\boldsymbol{\beta}_{s+1,n}) & \text{otherwise} \end{bmatrix}, \\
\Sigma_{r,s} &= \tilde{K}_{r,s}^{-1},
\end{aligned}
$$

which was calculated using standard properties of the multivariate Gaussian distribution. In this calculation the precision matrix is decomposed into

$$K = \left(\begin{array}{ccc} & \tilde{K}^T_{-p+1,r-1} & \\ \tilde{K}_{-p+1,r-1} & \tilde{K}_{r,s} & \tilde{K}_{s+1,n} \\ & \tilde{K}^T_{s+1,n} & \end{array} \right),$$

where $\tilde{K}_{r,s}$ is the square $gq \times gq$ matrix containing blocks $K_{r,r}$ to $K_{s,s}$. The remaining two blocks are rectangular, contain the same rows as $\tilde{K}_{r,s}$, and include all the remaining columns. To avoid any mixing problems at the boundaries of each block, the length of the first block can be randomly generated from the set $\{q, 2q, \ldots, gq\}$. The acceptance probability of a move from $\boldsymbol{\beta}^{j-1}_{r,s}$ to $\boldsymbol{\beta}^{*}_{r,s}$ is given by

$$\min \left\{ 1, \frac{\prod_{t=r}^{s} \mathrm{Poisson}(y_t | \boldsymbol{\beta}^{*}_{t}, \boldsymbol{\alpha}^{j-1})}{\prod_{t=r}^{s} \mathrm{Poisson}(y_t | \boldsymbol{\beta}^{j-1}_{t}, \boldsymbol{\alpha}^{j-1})} \right\}.$$

Further details can be found in Knorr-Held (1999).

References

Ameen, J. and P. Harrison (1985). Normal Discount Bayesian Models. *Bayesian Statistics 2 2*, 271–198.

Chatfield, C. (1996). *The Analysis of Time Series: An Introduction* (5th ed.). Chapman and Hall.

Chiogna, M. and C. Gaetan (2002). Dynamic generalized linear models with applications to environmental epidemiology. *Applied Statistics 51*, 453–468.

Daniels, M., F. Dominici, S. Zeger, and J. Samet (2004). The National Morbidity, Mortality, and Air Pollution Study Part III: Concentration-Response Curves and Thresholds for the 20 Largest US Cities. *HEI Project 96-97*, 1–21.

Dominici, F., M. Daniels, S. Zeger, and J. Samet (2002). Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association 97*, 100–111.

Dominici, F., J. Samet, and S. Zeger (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: a hierarchical modelling strategy. *Journal of the Royal Statistical Society series A 163*, 263–302.

Fahrmeir, L. (1992). Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models. *Journal of the American Statistical Association 87*, 501–509.

Fahrmeir, L., W. Hennevogl, and K. Klemme (1992). Smoothing in dynamic generalized linear models by Gibbs sampling. *Advances in GLIM and Statistical Modelling*, 85–90.

Fahrmeir, L. and H. Kaufmann (1991). On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression. *Metrika 38*, 37–60.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.). Springer.

Fahrmeir, L. and S. Wagenpfeil (1997). Penalized likelihood estimation and iterative Kalman smoothing for non-Gaussian dynamic regression models. *Computational Statistics and Data Analysis 24*, 295–320.

Fruhwirth-Schnatter, S. (1994). Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering. *Statistics and Computing 4*, 259–269.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika 85*, 215–227.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2003). *Bayesian Data Analysis* (2nd ed.). Chapman and Hall.

Griffin, M. and K. Neuzil (2002). The Global Implications of Influenza in Hong Kong. *The New England Journal Of Medicine 347*, 2159–2162.

Katsouyanni, K., J. Schwartz, C. Spix, G. Touloumi, D. Zmirou, A. Zanobetti, B. Wojtyniak, J. Vonk, A. Tobias, A. Ponka, S. Medina, L. Bacharove, and H. Anderson (1996). Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. *Journal of Epidemiology and Community Health 50*, S12–S18.

Kitagawa, G. (1987). Non-Gaussian State-Space Modelling of Nonstationary Time Series. *Journal of the American Statistical Association 82*, 1032–1041.

Kitagawa, G. (1996). Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-Space Models. *Journal of Computational and Graphical Statistics 5*, 1–25.

Knorr-Held, L. (1999). Conditional Prior Proposals in Dynamic Models. *Scandinavian Journal of Statistics 26*, 129–144.

Peng, R., F. Dominici, R. Pastor-Barriuso, S. Zeger, and J. Samet (2005). Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities. *American Journal of Epidemiology 161*, 585–594.

Pettit, L. (1986). Diagnostics in Bayesian model choice. *The Statistician 35*, 183–190.

Roberts, S. (2004). Biologically Plausible Particulate Air Pollution Mortality Concentration-Response Functions. *Environmental Health Perspectives 112*, 309–313.

Samet, J., S. Zeger, F. Dominici, F. Curriero, I. Coursac, D. Dockery, J. Schwartz, and A. Zanobetti (2000). The National Morbidity, Mortality, and Air Pollution Study Part II: Morbidity and Mortality, from Air Pollution in the United States. *HEI Project 96-97*, 5–47.

Shaddick, G. and J. Wakefield (2002). Modelling multiple pollutants and multiple sites. *Applied Statistics 51*, 351–372.

Shephard, N. and M. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Spiegelalter, D., N. Best, B. Carlin, and A. Van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society series B 64*, 583–639.

Vedal, S., M. Brauer, R. White, and J. Petkau (2003). Air Pollution and Daily Mortality in a City with Low Levels of Pollution. *Environmental Health Perpespectives 111*, 45–51.

West, M., J. Harrison, and H. Migon (1985). Dynamic Generalized Linear Models and Bayesian Forecasting. *Journal of the American Statistical Association 80*, 73–83.

Zhu, L., B. Carlin, and A. Gelfand (2003). Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics 14*, 537–557.

Zmirou, D., J. Schwartz, M. Saez, A. Zanobetti, B. Wojtyniak, G. Touloumi, C. Spix, A. Ponce de Leon, Y. Le Moullec, L. Bacharova, J. Schouten, A. Ponka, and K. Katsouyanni (1998). Time-Series Analysis of Air Pollution and Cause Specific Mortality. *Epidemiology 9*, 495–503.

Table 1: Summary of the eight models. The base model is given by (5).

Model	**Trend** β_t	**Air pollution effect** γ_t
1	(a) - splines	(i) - constant
2	(a) - splines	(ii) - random walk
3	(b) - first order random walk	(i) - constant
4	(b) - first order random walk	(ii) - random walk
5	(c) - second order random walk	(i) - constant
6	(c) - second order random walk	(ii) - random walk
7	(d) - local linear trend	(i) - constant
8	(d) - local linear trend	(ii) - random walk

Table 2: Summary of the smoothing parameters from the Bayesian and likelihood analyses.

Model	**Parameter**	**Bayesian**			**Likelihood**
		2.5%	median	97.5%	
1	k	*NA*	27	*NA*	27
2	k	*NA*	27	*NA*	27
	σ^2	9.91×10^{-8}	1.10×10^{-7}	1.21×10^{-7}	0
3	τ^2	0.00018	0.00019	0.00021	0.373
4	τ^2	0.00018	0.00019	0.00021	0.373
	σ^2	1.01×10^{-7}	1.10×10^{-7}	1.20×10^{-7}	0
5	τ^2	2.20×10^{-6}	3.56×10^{-6}	5.78×10^{-6}	0.004
6	τ^2	2.27×10^{-6}	3.67×10^{-6}	5.95×10^{-6}	0.004
	σ^2	1.19×10^{-7}	8.25×10^{-7}	1.66×10^{-6}	0
7	τ^2	1.61×10^{-7}	1.68×10^{-7}	1.76×10^{-7}	10^{-7}
	ψ^2	3.00×10^{-6}	3.09×10^{-6}	3.18×10^{-6}	0.003
8	τ^2	1.58×10^{-7}	1.67×10^{-7}	1.77×10^{-7}	10^{-7}
	ψ^2	3.05×10^{-6}	3.17×10^{-6}	3.25×10^{-6}	0.003
	σ^2	7.43×10^{-7}	2.64×10^{-6}	5.44×10^{-6}	0

Table 3: Relative risks for an increase in $10\mu g/m^3$ of PM_{10} and corresponding 95% credible or confidence intervals.

Model	**Bayesian**	**Likelihood**
1	1.007 (0.998 , 1.016)	1.015 (1.010 , 1.020)
3	1.011 (1.002 , 1.020)	1.014 (1.008 , 1.019)
5	1.008 (0.999 , 1.017)	1.014 (1.008 , 1.019)
7	1.009 (1.000 , 1.019)	1.013 (1.007 , 1.018)

Figure 1: The health data and the estimated trends from the four approaches: (a) natural cubic spline; (b) first order random walk; (c) second order random walk; (d) local linear trend. Bayesian and likelihood estimates are represented by solid and dashed lines respectively.

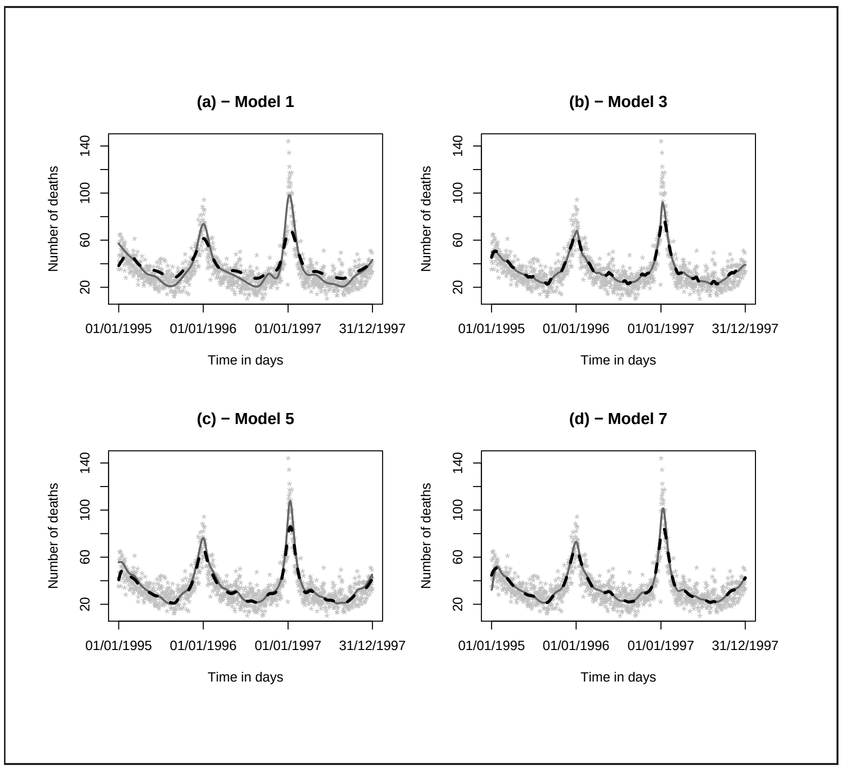

Figure 2: Autocorrelation function of the Bayesian residuals for different type of trends: (a) natural cubic spline; (b) first order random walk; (c) second order random walk; (d) local linear trend.

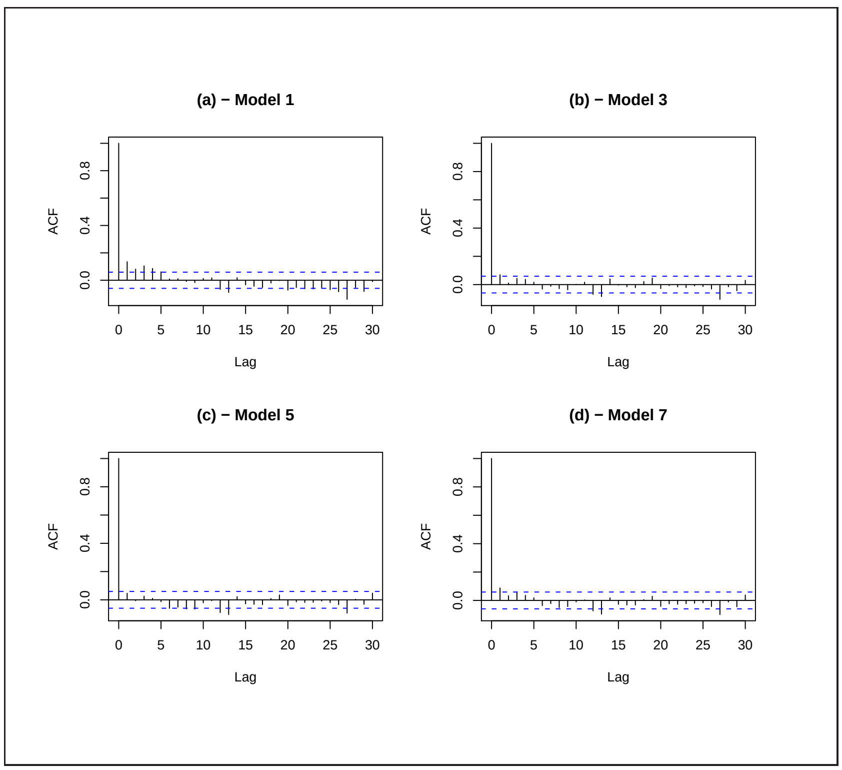

Figure 3: Estimated time-varying effects (solid line) of PM_{10} from the Bayesian analyses with 95% credible intervals. The dotted line represents a constant effect over time. The four panels relate to the different trend models: (a) natural cubic spline; (b) first order random walk; (c) second order random walk; (d) local linear trend.